Brazilian Journal of Biometrics, **43**, 1–11 (2025)
doi:10.28951/bjb.v43i3.773

ARTICLE

Probabilistic model based on the Skellam distribution: application in sports betting.

Alexsandro Andre Alves de Freitas*[,1] and Alessandra dos Santos[2]

[1]Undergraduate at Center of Exact and Technological Sciences, Western Paraná State University, Cascavel, PR, Brazil
[2]Professor at Center of Exact and Technological Sciences, Western Paraná State University, Cascavel, PR, Brazil
*Corresponding author. Email: alexsandroandre050@gmail.com

(Received: September 26, 2024; Revised: April 8, 2025; Accepted: May 5, 2025; Published: June 24, 2025)

Abstract

Probabilistic models are widely used in various areas of knowledge to estimate the chance of random events occurring. In Brazil, the sports market has adopted this tool to assist in decisions such as investing in certain specificities of the subject, evaluating accuracy of players and predicting actions of opponents. Furthermore, this market has grown with the emergence of several betting houses, specially due to dissemination in the open TV channels. This work focused on exploring Skellam's probabilistic models, analyzing their principles and characteristics to model the men's Brazilian soccer championship called Brasileirão. The database includes historical match results from 2012 to 2023 and estimates of the odds provided by the Bet365 bookmaker. Three probabilistic models were proposed to calculate the probabilities outcomes (win, loss or draw) and, consequently, the odds in each game; these results were compared with those of Bet365 and it showed efficiency. However, it is important to emphasize that the models have limitations due to external factors that have not been evaluated, such as weather conditions and game strategies.

Keywords: Brazilian soccer championship; Odds; Profit; Stakes.

1. Introduction

Sports betting is a popular form of entertainment that allows enthusiasts to become more involved with their favorite teams and competitions. According to Chagas (2016), the technology and the vast amount of information have transformed this pastime into something sophisticated, where decision-making is crucial to success. Thus, probabilistic models are fundamental, providing quantitative approaches to evaluate probabilities of outcomes.

Some studies are based on statistical models to predict results of games. For example, Goddard & Asimakopoulos (2004) used an ordered probit regression model to forecast English league football results, incorporating various explanatory variables such as cup competition involvement and geographical distance between teams. They found that their model contains information about match outcomes not reflected in bookmakers' odds, indicating weak form inefficiency, particularly in matches towards the end of the season. Galvão & Bueno Filho (2020) fitted Poisson models with linear predictors to estimate the expected number of goals for home and away teams in the 2017 Brazilian Championship. They concluded that home advantage can be worked out conditioning on point estimates and works really well to enhance predictions. Similarly, Robbins (2023) concluded that while the sports betting market is generally efficient, sportsbooks have several advantages, including offering unfair bets that create a margin for profit regardless of the outcome, as long as bets are placed in the appropriate proportion. This demonstrates the potential of statistical models to improve sports betting strategies.

Furthermore, Newall & Cortis (2021) suggest that sports bettors may be biased toward both longshots and genuine favorites. Their review indicates that biases exist, but further investigation and larger datasets are needed to understand these phenomena fully. Hegarty & Whelan (2024) found that fixed-odds betting markets for soccer and tennis do not satisfy the strong form definition of market efficiency. Their analysis revealed a substantial favorite-longshot bias, with higher expected returns on favorites compared to longshots. However, this bias does not translate into profits for bettors, supporting weak-form market efficiency. They also highlighted the importance of using normalized probabilities in regressions testing market efficiency.

Additionally, Newall *et al.* (2021) showed that skill can lead to positive long-run returns in certain sports betting situations. However, the structural characteristics of modern sports betting products often lead bettors to make bets with long-run losses due to high betting frequency, low win probability, and high involvement. In our analysis, we address the important issue of staking methods in sports betting. Barge-Gil & Garcia-Hiernaux (2020) discussed the Kelly staking criterion, which maximizes long-term growth for expected positive returns but requires exact estimation of true probabilities. They suggest that fractional Kelly or shrinkage factors should be applied to account for uncertainty. However, many bettors opt for simpler strategies like the unit-loss or unit-win strategies.

To enhance prediction results, we propose the Skellam model, which is particularly suitable for modeling score differences in sports events where the outcome can be described by two independent Poisson-distributed random variables. This model can provide a more accurate assessment of the probabilities of different match outcomes, thereby improving betting strategies.

Guimarães (2022) affirms that around 97% of bettors lose money in the long term, highlighting the importance of methods based on mathematical models. This study aims to understand the Skellam model and evaluate its effectiveness in predicting sports betting outcomes, ultimately seeking to improve long-term profitability for bettors.

The Brasileirão Championship (Campeonato Brasileiro Série A) features 20 teams competing over 19 rounds per season, and this analysis covers 12 seasons from 2012 to 2023. As Brazil's top-tier football league, founded in 1959, it follows a round-robin format, with the top teams qualifying for the Copa Libertadores and the four lowest-ranked teams relegated to Série B each year. The diversity in match schedules, played at different times of day and under various conditions, adds an intriguing element to performance analysis. This and other official information can be found on the *Brazilian Football Confederation (CBF)* website [1].

The structure of the article is as follows: the second section presents the database and methodology used, the third section reports the results achieved, and the fourth section provides some concluding remarks.

[1] https://www.cbf.com.br/

2. Materials and Methods

The database was obtained from the website *football-data*[2]. The set contains information from 4,522 championship matches, from 2012 to 2023. Every game presents: the season in which the match was played, the date of the match, the home team, the away team, the number of goals scored by the home team, the number of goals scored by the away team, the match winner, the pre-match home win odd, the pre-match draw odd, the pre-match away win odd, the maximum pre-match home win odd, the maximum pre-match draw odd, the maximum pre-match away win odd, the average pre-match home win odd, the average pre-match draw odd, and the average pre-match away win odd.

Initially, a descriptive analysis was performed using boxplot graphics, calculations of the average goals per team, and the absolute frequency of wins for each team, both when playing at home and away.

Based on the descriptive data, three models were developed using the Skellam distribution to analyze football matches. The model uses the average home goals, the average away goals, the number of home wins per season, and the number of away wins per season, considering a specific period of years from the database (2012 to 2022). These averages are used to calculate the probabilities of the possible goal differences (d) between teams, applying the Skellam distribution (Equation 1).

$$P(D = d) = e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{d}{2}} I_{|d|}(2\sqrt{\lambda_1\lambda_2}) \tag{1}$$

with $-10 \leq d \leq 11$, λ_1 being the average home goals, λ_2 being the average away goals and $I_{|d|}(2\sqrt{\lambda_1\lambda_2})$ being the modified Bessel function of the first kind of order $|\, d \,|$.

The calculated probabilities are grouped into three main events: home win ($i = 1$ and $P(D > 0)$), draw ($i = 2$ and $P(D = 0)$) and away win ($i = 3$ and $P(D < 0)$). To ensure that the probabilities sum to 1, they are normalized as shows the Equation 2.

$$P_i^* = \frac{P_i}{\sum_{i=1}^{3} P_i} \tag{2}$$

where P_i^* is the i-th normalized event probability, P_i is the i-th raw probability of each event, and $\sum_{i=1}^{3} P_i$ is the sum of the raw probabilities.

The normalized probabilities are transformed into odds ($1/P_i^*$) and organized in a data frame to facilitate the analysis of the results.

For the first proposed model, data from all seasons were used. In the second, only the four most recent seasons (2019 to 2022) were considered, assuming that older games may have little influence on teams' current performance. In the third model, in addition to considering only the last four seasons, higher weights were assigned to more recent matches using an exponential function to weight the date of j-th match, such as $\text{Weight}_j = e^{-\left(\frac{\text{Today's Date}-\text{Match Date}_j}{365}\right)}$. This was applied to the t-th team average goals (λ_t).

For all versions of the model, we considered a fictitious bettor who makes bets when the odd calculated by the model has a percentage difference of at least 20% compared to the Bet365 odds[3]. It was assumed a stake of R$1,000.00 per bet and the Return On Investment (ROI - Equation 3).

$$ROI = \frac{\text{Profit from bets}}{\text{Total stake}} \times 100 \tag{3}$$

[2]https://www.football-data.co.uk/

[3]https://www.bet365.com/#/AC/B1/C1/D1002/E88369731/G40/

To obtain the Bet365 odds for games that have already been played, it is necessary to use an external statistics website, such as Sofascore[4].

All analyses were done in Python (version 3.10) using Google Colab[5]. The datasets and scripts are available in https://github.com/aulexsandro/odds.camp.brasil.

3. Results and Discussion

In the Brasileirão championship, the team that played at home had better results than the visiting team. Figure 1 and 2 show the number of goals average and wins. The median of the average number of goals per team were 1.39 when its plays in home and 0.89 for visiting, the mean were 1.41 and 0.95, respectively. For the number of wins the results were 9 and 4 for median when teams plays in home and away, respectively; the mean were 8.90 and 4.45.

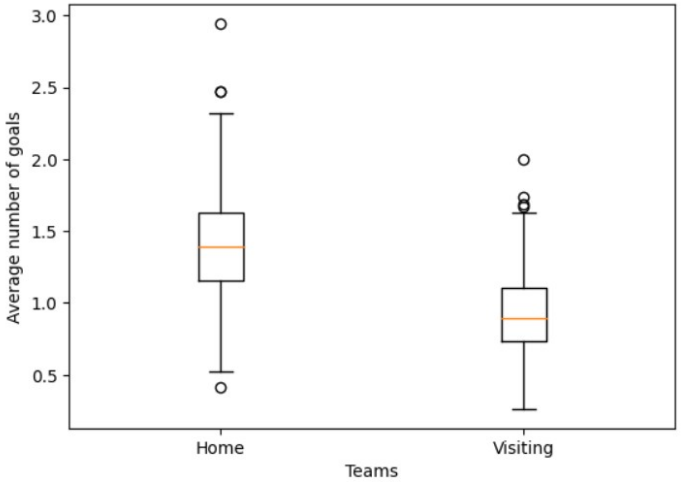

Figure 1. Boxplot of the average number of goals for each team according to the situation of playing at home or as visiting team, during 2012 to 2022.

[4]https://www.sofascore.com/tournament/football/brazil/brasileirao-serie-a/
[5]https://colab.research.google.com/

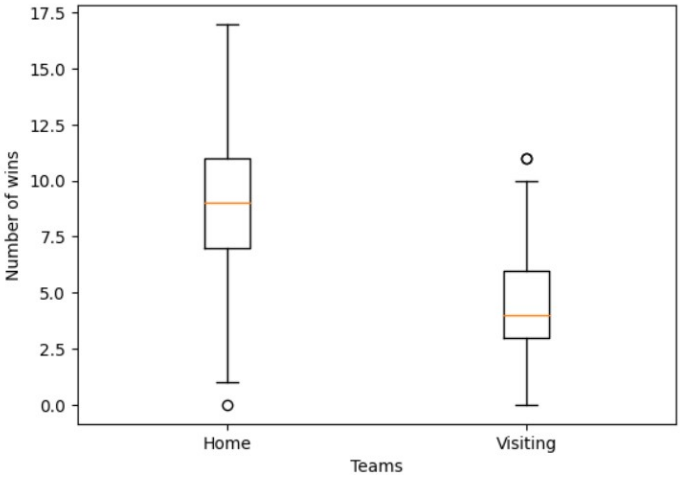

Figure 2. Boxplot of the wins number for each team according to the situation of playing at home or as visiting team, during 2012 to 2022.

For the first fitted model, the results compared with those of Bet365 resulted in 115 bets placed throughout the championship. Table 1 presents a summary of the number of bets and the profit (in percentage) in each round, including the cumulative profit at the last round. In the first round, the bettor had a profit of 288% on his stakes, but over the rounds he accumulated a loss of 958%, with an ROI of -8.33%. At the end, the total loss would have been R$9,580.00.

Table 1. Summary of bets placed for the first version of the fitted model for the 2023 championship

Round	Number of Bets	Round Result (%)	Cumulative Profit (%)
1	5	288	288
2	7	-120	168
3	8	143	311
4	5	325	636
5	7	-470	166
6	6	475	641
7	5	-225	416
8	7	275	691
9	7	-700	-9
10	5	-500	-509
11	7	110	-399
12	8	470	71
13	5	-500	-429
14	6	91	-338
15	5	-500	-838
16	6	-280	-1118
17	6	90	-1028
18	6	70	-958
19	4	0	-958

In the second fitted model, different bets were placed; in the total, it was 101 bets. Despite some rounds with negative results, the model showed an accumulated profit of 189% on the stakes, with

an ROI of 1.87%. With a stake of R$1,000.00 per bet, the total profit would have been R$1,890.00. Table 2 summarizes the bets placed with this version of the model.

Table 2. Summary of bets placed for the second version of the fitted model for the 2023 championship

Round	Number of Bets	Round Result (%)	Cumulative Profit (%)
1	5	-500	-500
2	3	-300	-800
3	6	343	-457
4	7	125	-332
5	7	-182	-514
6	5	-25	-539
7	6	425	-114
8	5	478	364
9	4	-400	-36
10	6	-190	-226
11	6	630	404
12	7	350	754
13	5	-125	629
14	4	-400	229
15	4	-400	-171
16	7	270	99
17	5	190	289
18	5	-100	189
19	4	0	189

For the third fitted model it was created heatmaps to show the percentage differences among the odds calculated by the model and those from Bet365. Figure 3 presents the results for the home teams win and Figure 4 for the visiting teams win, both for the first round. The same was made for the other rounds. Figure 5 presents a boxplot of the differences between the odds calculated by the third model and the Bet365 for all championship games in round one and two. The discrepancies highlight the model's overestimation of visiting teams and underestimation of home teams, in contrast with Bet365.

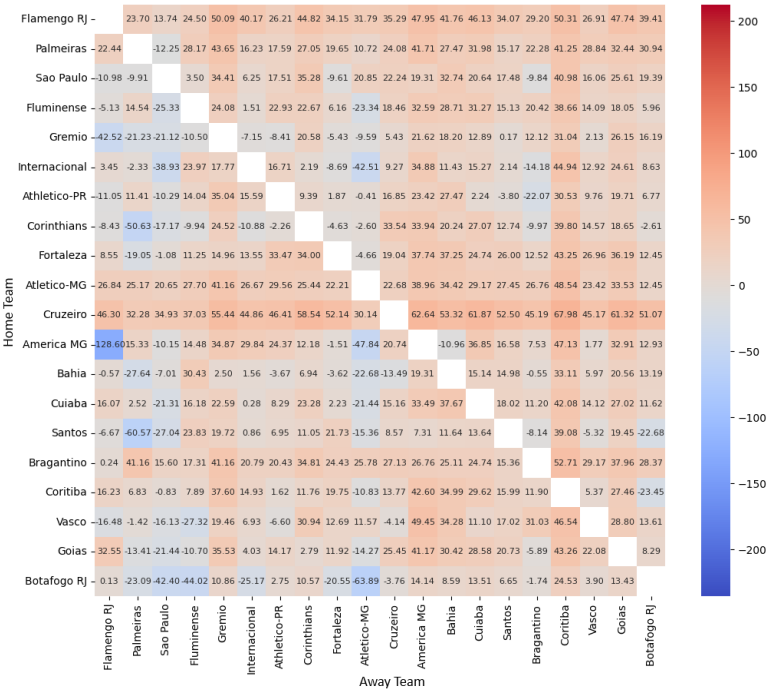

Figure 3. Percentage differences in the odds for home team wins in the first round of the 2023 championship.

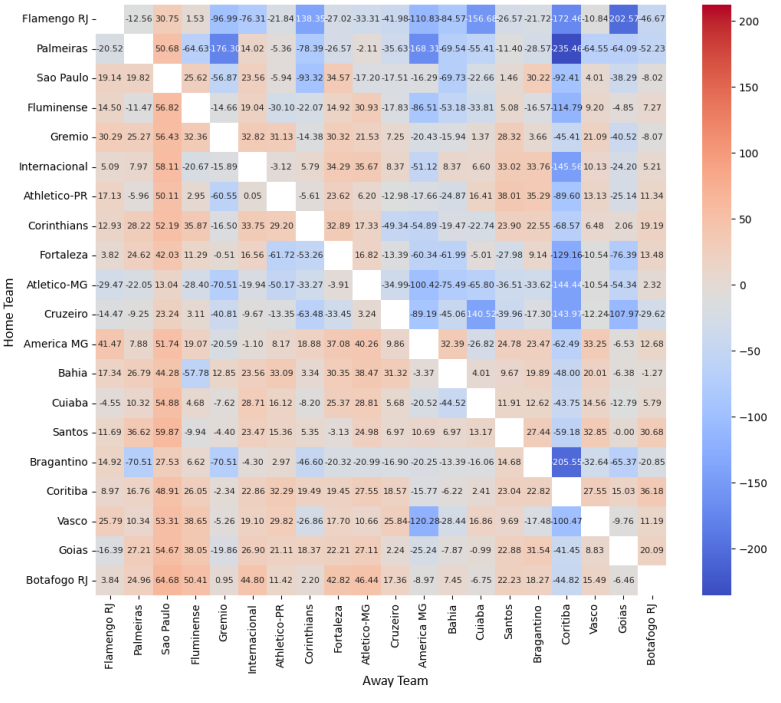

Figure 4. Percentage differences in the odds for visiting teams win in the first round of the 2023 championship.

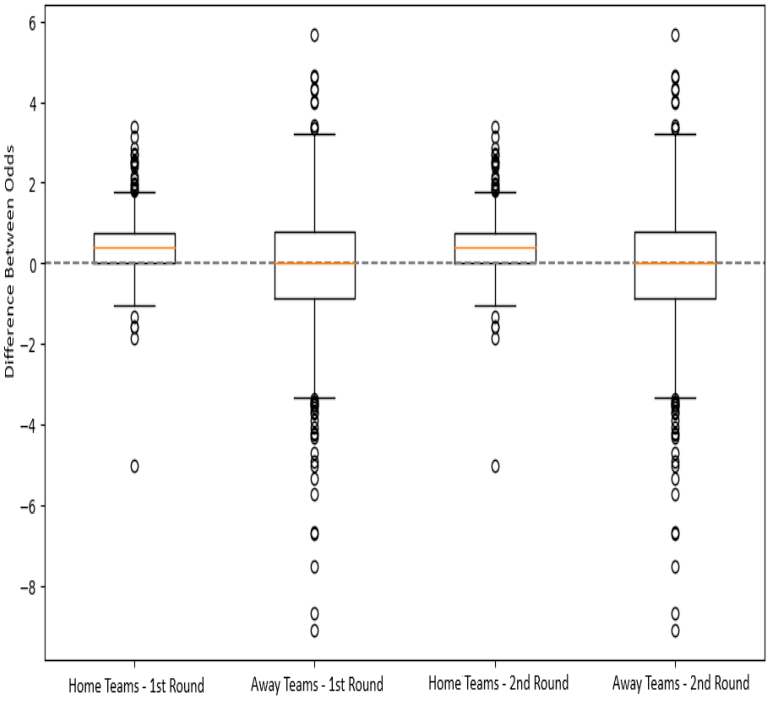

Figure 5. Boxplots of the differences between the odds from the third fitted model and Bet365 in the 2023 championship, when home or visiting teams win for first and second round.

Despite the discrepancies in some specific games, the model showed an accumulated profit of 557% on the stakes, with an ROI of 7.23%. The profit would have been R$5,570.00. Table 3 summarizes the results of this version, which suggested a total of 77 bets.

Table 3. Summary of bets placed for the third version of the fitted model for the 2023 championship

Round	Number of Bets	Round Result (%)	Cumulative Profit (%)
1	5	-262	-262
2	2	-200	-462
3	7	986	262
4	4	425	687
5	4	-400	287
6	5	-500	-213
7	4	625	412
8	6	-100	312
9	3	-90	222
10	4	-175	47
11	4	380	427
12	4	250	677
13	3	75	752
14	4	100	852
15	2	-200	652
16	5	425	1077
17	4	-50	1027
18	1	-100	927
19	6	-370	557

It is important to note that while the first version of the model resulted in a financial loss, the second and third versions showed profitable performance, suggesting that adjustments in calculation methodologies led to substantial improvements in the model's performance.

For more information and a detailed analysis, we recommend reading the complete work, which is available at de Freitas (2024).

4. Conclusions

This work analyzed the application of probability theory in sports betting, focusing on the men's Brazilian soccer championship. Probabilistic models were developed to calculate the odds of each match, demonstrating accuracy comparable to the odds provided by Bet365. The results showed that the model can predict match outcomes with good accuracy, capturing nuances and patterns from historical data.

Despite its effectiveness, the model faces limitations due to unpredictable external factors such as injuries, weather conditions and unexpected events during the games. However, it offers a data-driven approach, increasing the chances of success in betting and promoting a more systematic understanding of the sport.

The model is subject to continuous improvements, incorporating new variables and advanced statistical modeling methods. Additionally, it can be adapted to other leagues and sports, requiring only the corresponding database swap. In summary, this study aims to promote more informed and evidence-based practices in sports betting, encouraging a technical and scientific approach among bettors.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization: FREITAS, A. A. A. **Data curation**: FREITAS, A. A. A. **Formal analysis**: FREITAS, A. A. A. **Funding acquisition**: FREITAS, A. A. A. **Investigation**: FREITAS, A. A. A. **Methodology**: FREITAS, A. A. A.; SANTOS, A. **Project administration**: FREITAS, A. A. A.; SANTOS, A. **Software**: FREITAS, A. A. A. **Resources**: FREITAS, A. A. A.; SANTOS, A. **Supervision**: SANTOS, A. **Validation**: FREITAS, A. A. A.; SANTOS, A. **Visualization**: FREITAS, A. A. A.; SANTOS, A. **Writing – original draft**: FREITAS, A. A. A. **Writing – review and editing:** SANTOS, A.

References

1. Barge-Gil, A. & Garcia-Hiernaux, A. Staking in sports betting under unknown probabilities: Practical guide for profitable bettors. *Journal of sports economics* **21,** 593–609 (2020).
2. Chagas, J. M. *A (im) possibilidade de regulamentação das apostas esportivas no ordenamento jurídico brasileiro* TCC (Graduação em Direito) (Universidade Federal de Santa Catarina, Florianópolis, 2016), 88. https://core.ac.uk/download/pdf/43285438.pdf.
3. De Freitas, A. A. A. *Modelo probabilístico para auxílio em apostas esportivas* TCC (Graduação em Licenciatura em Matemática) (Universidade Estadual do Oeste do Paraná, Cascavel, 2024), 115.
4. Galvão, L. R. & Bueno Filho, J. S. S. Estimação de rating no futebol: campeonato brasileiro de 2017. *Brazilian Journal of Biometrics* **38,** 1–17 (2020).
5. Goddard, J. & Asimakopoulos, I. Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting* **23,** 51–66 (2004).
6. Guimarães, G. *Por que perdi dinheiro apostando?* https://clubedaposta.com/punting/ganhar-dinheiro-apostas-esportivas/.
7. Hegarty, T. & Whelan, K. *Comparing two methods for testing the efficiency of sports betting markets* tech. rep. (UCD Centre for Economic Research Working Paper Series, 2024).

8. Newall, P. W. & Cortis, D. Are sports bettors biased toward longshots, favorites, or both? A literature review. *Risks* **9,** 22 (2021).
9. Newall, P. W., Russell, A. M. & Hing, N. Structural characteristics of fixed-odds sports betting products. *Journal of Behavioral Addictions* **10,** 371–380 (2021).
10. Robbins, T. R. Weak Form Efficiency in Sports Betting Markets. *American Journal of Management* **23** (2023).